

The Lending Club Case Study

Welcome to the lending club case study presentation

The Approach

While working on the data analysis for the Lending Club, I followed the following approach

-  **Identifying Variables** - Group and identify the variables that will help in risk analysis
-  **Data Cleaning And Preparation** - Clean the available Data and prepare it for analysis
-  **Univariate Analysis and Segmentation** - Conduct Univariate analysis and segmentation on the filtered Data
-  **Bivariate Analysis** - See the impact of two variables combined on the overall risk
-  **Derive Final Metrics** - Try and derive final metrics from the data and analyse high risk accounts
-  **Correlation Analysis** - Analyse the correlation between the variables
-  **Recommendations** - Recommendations based on the analysis

Data Cleaning and Manipulation

Aiming to clean and manipulate the data to finally reach a data form where we can work with the data. Will include the following steps while cleaning and manipulating data.

1. Find the number of columns which do not have any data.
2. Drop unwanted and duplicate rows.
3. Dropping text/description columns which will not contribute to overall analysis
4. Dropping column sub_grade as the current analysis will limit to Grade only
5. Dropping all columns which refer to behavioural data of customer post loan approval

Data Cleaning and Manipulation

6. Dropping columns which is unique id in nature. They dont contribute to loan analysis
7. Dropping all columns where all the values are NA
8. Dropping all columns with all zero values
9. Drop all columns who have constant values (ignoring NA value)
10. Drop the column if the amount of empty values is more than 65%

Data Conversion

Since we have removed unwanted data from the dataset now we can focus on data conversion. During this process we will focus on the following:

1. Converting columns to floats where needed.
2. Converting required columns to integers if they can be converted and exist as strings in dataset.
3. Round off any values in the columns which have floating point values with more than 2 decimal places.
4. Convert the columns loan_amnt and funded_amnt as float64

Data Conversion

5. Convert the term column into an integer from a string
6. Convert int_rate to float by removing the "%" character
7. Round off the values of key float fields to 2 decimal place
8. Converting the loan_status to boolean column. "Fully-Paid is True and Charged Off is False"
9. Converting the column issue_d from string object to DateTime

Univariate Analysis:

Customer Demographics:

- The majority of loan applicants have an annual income between 0 to 40K.
- The majority of the debt-to-income ratio falls within 0 to 20, with some going up to 30.
- The majority of applicants are either renting or have a mortgage.
- The highest number of loan applications are for debt consolidation purposes.
- California state has the highest number of loan applications.
- Most loan applicants have no public record of bankruptcy.
- The majority of customers have been employed for 10+ years or between 0-2 years.

Univariate Analysis:

Loan Demographics:

- The highest number of loan applications fall within the range of 5k to 10k.
- The majority of interest rates are between 5% to 16%, with some going up to 22%.
- Most installment amounts fall within the range of 20.
- The majority of loan applications are for a term of 36 months.
- Most loan application counts fall under the category of Grade B.

Univariate Analysis:

Time-Based Analysis:

- Loan application counts are increasing each year, with the highest volume of loan applications in Quarter 4 of every year.
- The lowest loan applications are in Q1, possibly due to financial challenges, festive seasons, or debt consolidation at the end of the year.

Univariate Analysis:

Inferences:

- The customer demographic data indicates which segment of customers to target for the highest volume of loans.
- Further analysis is needed to understand why other categories are not as high as a few.
- The LendingClub should be prepared for the highest volume of loans in Q4 and target customers in other quarters to increase sales.

Final Recommendations Based On Complete Process

- Borrowers who take loans for a term of 60 months are more likely to default.
- Borrowers with less than 1 year of employment have a higher risk on defaulting.
- Borrowers with a "Verified" loan status who take high loan amounts with a 60-month tenure have a higher risk of defaulting.
- Borrowers who have "Rent" and "Mortgage" as their home_ownership have a higher probability of defaulting.
- Borrowers who take loans for the purpose of debt consolidation are at a higher risk of defaulting.
- Borrowers who take loans for the purpose of small business have a higher probability of defaulting.

Final Recommendations Based On Complete Process

- Borrowers with loan amount 15K and above is at the highest risk.
- Borrowers who receive interest rates between 15% and higher are more likely to default.
- Borrowers with lower grades, i.e., E,F,G, are at a higher risk of defaulting.
- Borrowers from NV (Nevada) have a high risk of Charge Offs.
- Borrowers having bankruptcy record are at high risk of Charge Offs
- Borrowers with pub_rec_bankruptcies count of 2 and higher have even higher Charge Off ratio
- Borrowers with income range of 0-40K have the highest risk of Charge Offs